DOCUMENT RESUME

ED 141 965 EC 101 178

AUTHOR Evans, Robert A.; Wozney, Barbara E.
TITLE Intervention Effect on LD Achievement.
PUB DATE Apr 77
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, New York, April, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Effective Teaching; *Elementary Secondary Education; Intervention; *Learning Disabilities; Program Effectiveness; Research Projects; *Resource Centers; *Special Classes; Standardized Tests

ABSTRACT

A norm referenced model was used to evaluate the relative effectiveness of intervention efforts with 3,896 learning disabled elementary and secondary school children. Ss were tested on the Wide Range Achievement Test. Significant growth was observed in students from resource and self contained classrooms. Gains did not appear to vary with the intensity of the intervention program. (GW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141965

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

INTERVENTION EFFECT ON LD ACHIEVEMENT*

by

Robert A. Evans
Evaluation Specialist
and
Barbara E. Wozney
Learning Disabilities Program Specialist
Division of Special Education
Fairfax County Public Schools
Fairfax, Virginia

EC 101178

* Paper presented to the American Educational Research Association at the April 1977 annual meeting in New York City, Session 8.05.

INTRODUCTION

Educational programs, particularly special education, continually require feedback on progress. Program planning as well as staff morale dictate this need. Ideally, the evaluation team takes part in the planning effort from a program's inception. Typically, this not the case, with evaluators being called upon to assess an on-going program. An attempt was made under these latter circumstances to evaluate a learning disabilities (LD) program. The effort presented here is viewed as an initial attempt not only to document progress but to also assess a particular design.

Intervention Effect

Researchers functioning within public educational settings have frequently been plagued by an inability to generate randomized experimental models. This is particularly evident with respect to special education programs where legal and ethical restrictions prevent the establishment of true control groups as well as randomized instructional or treatment assignments.

One strategy which attempts to compensate for these restrictions is the establishment of quasi-experimental models suggested by Campbell and Stanly (1963). These procedures, however, are still considered unacceptable in many educational research environments. The problem of non-randomized treatment groups remains to exert its heavy influence, often resulting in conclusions of questionable merit.

An alternative analysis approach for program evaluation is presented within this paper. Instructional treatments are primarily aimed at generating change or growth, and change can be thought of in two ways: actual or observed, and expected. Actual or observed change can be directly compared with expected change resulting in what can be termed as intervention effect (i.e., the change attributable to the educational treatment). In assessing this effect, a

pre-test and post-test are required. Post-test scores comprise the observed or actual end of program results. Each student's pre-test provides the basis for projecting expected results. The percentile or standard score position within the normative data (supplied by the test manual) is not anticipated to change in the absence of a treatment effect. Therefore, if the educational program did not result in improved skill acquisition as measured by test performance, percentile or standard score position would remain statistically the same.

Rationale

The Wide Range Achievement Test (WRAT) was selected as the evaluation instrument. Economy of time and resources were considered as critical factors in this initial effort. Cooperation by the classroom teacher was the prime ingredient in making the evaluation process a success.

In formulating the hypotheses, generally held beliefs were to be verified. Students assigned to the resource model indicated a need for a restricted level of treatment. Conversely, the greater demonstrated needs of the more severe handicapped LD students dictated their placement in a self-contained setting. If this assumption was valid, then the results should indicate pre-test differences, with lower scores generated by the self-contained group. Rejection of H_2 would reinforce the notion of group differences.

Greater gain in achievement scores have traditionally been associated with the intensity of service provided with little thought given to degree of handicapping condition. The third hypothesis was formulated to test this concept.

METHOD

Subjects

The LD program of Fairfax County Public Schools (Virginia) provided services for 3,896 children at various levels and within differing configurations for the academic year 1975-1976. Self-contained and resource classes served 2,744 students at elementary level. At intermediate and secondary levels a resource model only was used to provide services to 1,152 students. Fairfax County is divided into four administrative areas, with LD programs of each model (resource; self-contained), operating in each area.

Resource models provided services to 3,558 handicapped children for as much daily time as was feasible and appropriate with a pronounced emphasis on mainstreaming. The level of participation was a function of the child's needs. Therefore, more severely handicapped children received more intense levels of service, while children with less urgent needs participated at less intense levels.

Children whose needs indicated a requirement for an intensive level of service, were enrolled in self-contained LD classes on a full-time basis and accounted for 334 children: Even at this level, however, mainstreaming, where feasible and appropriate, was instituted for these children.

Hypotheses

H_1: There are no differences between WRAT pre- and post-test results for the LD program, for both resource and self-contained models.

H_2: There are no differences between resource and self-contained pre-test WRAT results.

H_3: There are no differences in achievement gains between resource and self-contained models.

H_4: There are no differences between observed and expected post-test results (intervention effects) for the LD program.

H_5: There are no differences in the intervention effect for the resource and self-contained models.

Statistics

A full regression analysis was performed using the elementary sample of 2,744. Program models (resource; self-contained), area membership and sex were included as predictor variables, with WRAT pre-test results as dependent or criterion variables.

An analysis of co-variance was employed in determining the intervention effect because pre-test differences existed between the various program models. The pre-test was used as a co-variate to compensate for initial differences.

The regression analysis consisted of data from elementary, intermediate, and secondary students with pro-rata post-test scores as dependent variables.

Finally, the t test for correlated means was used to test the difference between pre- and post-tests, both expected and observed. The formula was as follows:

$$t_{N-1} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2 - 2r_{12} S_1 S_2}{N-1}}}$$

where:

$\bar{X}_1$ - observed mean post-test score

$\bar{X}_2$ - expected mean post-test score

S_1 - pre-test standard deviation

S_2 - post-test standard deviation

r_{12} - correlation between pre- and post-test scores

N - Number in sample

N - 1 - degrees of freedom

The above formula was adapted for use in analyzing observed pre- and post-test results. $\overline{X}_1$ and $\overline{X}_2$ then become observed mean pre- and post-test respectively.

Foat, Horst, Piestrap, and Tallmadge suggest the "educational significance" can be accepted where gains are observed to be at least one-third a standard deviation greater than expectations based on norm or control group scores (1976, p. 1). Horst, Tallmadge, and Wood (1975) supplied the following formula to measure educational significance:

$$\overline{Y} - \overline{\overline{Y}} = \sigma/3$$

where:

$\overline{Y}$ - observed mean post-test score

$\overline{\overline{Y}}$ - expected mean post-test score

$\sigma/3$ - one-third standard deviation.

Procedure

The WRAT was administered to all the students enrolled in the LD program using a pre-/post-testing paradigm. The average length of time between test administrations was 195 days or 6.5 months. The standard deviation for this time interval was 49 days. Therefore, 68 percent of the students were tested at intervals ranging between five and eight months apart. The pre-test administration cut-off date was January, 1976. Less than one percent or 38 children in the sample had to be dropped from the analysis as a result of incomplete data. Data supplied by special education teachers were verified for completeness and accuracy.

Transformation to standard scores and grade equivalents was derived from raw score data. Equalization of time intervals between pre- and post-tests was completed because of some observed variance in test administration between individual teachers. This time interval became important in analyzing educational significance. The linear pro-rata formula below was applied to allow group gains to be analyzed.

$$\text{Pro-rate gain} = \frac{X}{Y}(Z_1 - Z_2)$$

where:

X - actual time in school year (including weekends and holidays)

Y - actual time interval between test administrations

Z_1 - post-test score

Z_2 - pre-test score

Pro-rata post-test = pro-rata gain + pre-test

The expected (or no treatment) estimate was derived using the norm-referenced model presented in A Practical Guide to Measuring Project Impact on Student Achievement (Horst, Tallmadge, and Wood, 1975). Standard score means and standard deviations were computed for both pre- and post-test distributions, along with the correlation between these data. The percentile equivalents for these means were identified from appropriate norm tables. Using the pre-test percentile, the no treatment post-test expectation was derived. Therefore, absence of any special treatment would result in the group maintaining its relative position within the norm group. Mean age at post-test and pre-test percentile position yielded the expected mean post-test score for the group. The observed or actual post-test mean was examined in relation to the expected mean. Intervention effect became the difference between these means, with the observed being greater demonstrating a treatment effect. The statistical significance was determined using the t test for correlated means.

RESULTS

WRAT post-tests were significantly higher than pre-tests for the entire sample of 3,896 students in all three sub-test areas (reading, spelling, and arithmetic). The grade equivalent score was considered too unstable and inaccurate upon which to base conclusions, and for this reason that analysis was computed on standard scores as well as equivalent scores. The significant differences occurred for both observed or actual gains and pro-rata gains in the intervention effect on both measurement scales. Table I shows the standard scores and grade equivalent means, standard deviations, and values.

Table I

Gains in Achievement

Academic Area	Mean G.E.	S.D. G.E.	Mean Std. Score Gain	S.D. Std. Score	t Value G.E.	Intervention Effect G.E.	t Value Intervention Effect Std. Scores
Reading	1.4	1.8	5.9	18.3	47.85	26.21	20.12
Spelling	1.0	1.5	4.5	16.4	43.48	19.37	17.02
Arithmetic	1.3	1.9	6.1	20.0	41.86	20.11	19.04

The critical value for *t* at the .001 level of significance was 3.291. The results in Table I indicated *t* values in excess of the critical value for each academic area. Significant gains in achievement for the total sample were observed. Therefore, LD students as a group have made significant gains in achievement.

The elementary LD program models (self-contained; resource) were analyzed (N = 2,744). Table II shows the results of the pre-test analysis for models within this level. A full regression model with pre-test reading, spelling, and arithmetic as the dependent or criterion variables was used.

Predictor variables included program model, area membership and sex, with age included as a control because of the significant positive correlation between age and the pre-test levels (r = .65).

Table II

Regression Analysis Results for Pre-Tests with Elementary Resource and Self-Contained

	Reading	Spelling	Arithmetic
F-Ratio	67.9	66.02	61.0
DF 1	1	1	1
DF 2	2,735	2,735	2,735
Level of Significance	0,000	0,000	0,000

Table II, which shows results using grade equivalent data, indicates that pre-test scores for elementary resource students were significantly higher than those for self-contained students. The mean pre-test scores were as shown in Table III below.

Table III

Mean and Standard Deviation Sub-Test Scores on Pre-Test

	Reading		Spelling		Arithmetic	
LD	Mean	S.D.	Mean	S.D.	Mean	S.D.
Resource	3.059	1.755	2.555	1.325	2.917	1.257
Self-Contained	2.711	1.450	2.308	1.093	2.774	1.059

The significantly higher pre-test scores for resource students were obtained despite controlling for age and sex differences.

The results indicated substantial pre-test differences, therefore, analysis of co-variance was used to analyze achievement gains for the self-contained and resource models. Table IV shows results of this analysis.

Table IV

Analysis of Co-Variance with Self-Contained

Resource Gains for Total Sample LD Students

	Reading	Spelling	Arithmetic
F-Ratio	0	.892	.329
DF 1	1	1	1
DF 2	3884	3884	3884
Level of Significance	P .05	P .05	P .05

Table IV indicates that there were no significant differences between LD resource and self-contained achievement gains. The procedure as described by Horst, Tallmadge, and Wood (1975) was used in analyzing the intervention effect for each program model. Pre-test standard scores were used to determine the expected post-test (no treatment) effect. Because standard scores on WRAT norm tables correspond to percentiles, rounded pre-test means became the expected no treatment post-test means. Table V shows t valuer for the significance tests of the difference between actual and expected no treatment post-test results for each program model.

Table V

t Values by Academic Area for Observed

and Expected Post-Test Results by Program in Standard Scores

	RESOURCE				SELF-CONTAINED			
Academic Area	Actual Mean Post-Test	Expected No Treat. Post-Test	Diff.	t Value	Actual Mean Post-Test	Expected No Treat. Post-Test	Diff.	t Value
Read.	92.13	88	4.13	29.00	85.93	82	3.93	10.62
Spell.	85.36	83	2.36	16.86	81.65	78	3.65	9.60
Arith.	89.67	86	3.67	12.66	87.00	82	5.00	8.62

Each *t* value exceeds the .001 critical value of 3.291. Therefore, significant achievement gains were made by the LD students in both the resource and self-contained model.

Applying the test for educational significance to the above results, one-third standard deviation ($\sigma/3$) for the WRAT is(15/3) or 5. Therefore; only the spelling results for the self-contained group could be considered educationally significant.

DISCUSSION

The results of this initial program evaluation indicated the failure to accept all the posed hypotheses, except H_3. Achievement gains were equal for resource and self-contained students, that is, gains did not appear to be a function of intensity of service. LD students as a whole demonstrated gains in achievement as measured by the WRAT. These gains were observed using both actual and expected post-test results.

While Horst, Tallmadge, and Wood (1975) did not employ multiple *t* tests, their utilization here was primarily for demonstrating the utility of the approach. However, despite the application of the multiple *t* approach, the probability of a Type I error (rejecting a true null hypothesis) is only 19 percent.

The criteria established for resource and self-contained placement did differentiate between groups of students. Resource students did score higher than self-contained students.

It appears that LD students within the Fairfax County Public School system are being placed in a service model commensurate with their needs.

The use of the WRAT provided an economical and convenient measure of achievement for this population, although technical difficulties do seem to exist as Merwin and Thorndike (1972) have indicated. However, a concern for

the possibility of poor reliability due to low scores, restricted range of sample, etc may be unfounded. Julian Stanly indicated a low level of reliability "can be tolerated when . . . comparing mean differences" (p. 392). The critical statistics, according to Stanly is the standard error of measurement of the mean, not the reliability coefficient (1971).

While some evaluators may reject the use of pro-rata data as derived here, analysis of both pro-rata and non pro-rata reach significance. Pro-rata scores, however, may be used only for group data analysis, and not for individual students. This caution is dictated due to the large standard error of measurement on the WRAT.

The term "educational significance" may have greater meaning to administrators, teachers, and school boards than statistical significance. The introduction of terminology with which general educational personnel can utilize is viewed as an asset. The computation for educational significance, as depicted earlier, is simple and straight-forward. It is expected that its use will become more wide spread and generally accepted. Utilization of educational significance here was an effort to counter the effect of the large n. It is well recognized that small differences become significant as the number in the sample increases. Therefore, the results indicated that spelling was the only area where educational significance can be claimed.

CONCLUSION

This presentation demonstrated successful utilization of a norm-referenced model in educational evaluation. Care was taken to conserve local resources and effort for this initial attempt, leading to continued staff cooperation. Experimentation with other instruments is suggested, particularly those that have multiple data collection points represented in the standardization. It is believed that more precision can be achieved with instruments of

this kind.

Program evaluation, particularly in special education is a difficult task. Often evaluators are working within a restricted range of abilities. "True" no treatment control groups are essentially impossible for public school agencies. Legal and ethical restraints prohibit denying services to identified exceptional children. However, the intervention effect provides a useable and understandable model for assessment and dissemination of educational programs.

REFERENCES

Campbell, D. T., and Stanly, J. C., Experimental and qausi-experimental designs research, Chicago, Rand McNally, 1963.

Foat, C. M., Horst, D. P., Piestrap, A. M., and Tallmadge, G. K., The development of project information packages for the dissemination of effective educational programs. Paper presented to the American Educational Research Association at the annual meeting in San Francisco, California, April 19-23, 1976.

Horst, D. P., Tallmadge, G. K., and Wood, C. T., A practical guide to measuring project impact on student achievement, Washington, D. C.: U. S. Government Printing Office, 1975.

Merwin, J. C., Wide Range Achievement Test. In O. K. Buros (Ed.), The seventh mental measurements yearbook. (Vol. I). New Jersey: The Gryphon Press, 1972.

Stanly, J. C., Reliability. In R. L. Thorndike (Ed.) Educational measurement. Washington, D. C.: American Council on Education, 1971.

Tallmadge, G. K. and Horst, D. P., A procedural guide for validating achievement gains in educational projects. Washington, D. C. : U. S. Government Printing Office, 1976.

Thorndike, R. L., Wide Range Achievement Test. In O. K. Buros (Ed.) The seventh mental measurements yearbook (Vol. I). New Jersey: The Gryphon Press, 1972.